

General measure for macroscopic quantum states beyond “dead and alive”

Pavel Sekatski,¹ Benjamin Yadin,² Marc-Olivier Renou,³ Wolfgang Dür,¹ Nicolas Gisin,³ and Florian Fröwis³

¹*Institut für Theoretische Physik, Universität Innsbruck, Technikerstraße 21a, 6020 Innsbruck, Austria*

²*Department of Atomic and Laser Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*

³*Group of Applied Physics, University of Geneva, 1211 Geneva, Switzerland*

(Dated: April 10, 2017)

We consider the characterization of quantum superposition states beyond the pattern “dead and alive”. We propose a measure that is applicable to superpositions of multiple macroscopically distinct states, superpositions with different weights as well as mixed states. The measure is based on the mutual information to characterize the distinguishability between multiple superposition states. This allows us to overcome limitations of previous proposals, and to bridge the gap between general measures for macroscopic quantumness and measures for Schrödinger-cat type superpositions. We discuss a number of relevant examples, provide an alternative definition using basis-dependent quantum discord and reveal connections to other proposals in the literature. Finally, we also show the connection between the size of quantum states as quantified by our measure and their vulnerability to noise.

I. INTRODUCTION

When going from single microscopic particles to composite systems with many degrees of freedom, quantum mechanics shows enormous complexity. Genuine quantum features such as entanglement or nonlocality fall into several subclasses and notions such as “maximally entangled state” can not be generalized in a straightforward way. A full characterization of meso- or even macroscopic quantum systems seems to be out of reach, not only for practical reasons. However, one can identify global properties in such systems that are only mildly influenced by microscopic details. One such aspect is the macroscopic quantumness in large quantum systems.

An historic example that has played an important role is the so-called Schrödinger-cat state [1]. The pictorial idea of Schrödinger’s cat in a macroscopic superposition of dead and alive and entangled with a radioactive atom is easy to grasp. However, as first emphasized by Leggett [2], the so-called macroscopic distinctness of the two superposed components $|\text{alive}\rangle + |\text{dead}\rangle$ is a particularity not present in any quantum effect brought to macroscopic scales. For a counterexample, Leggett mentions superconductivity on visible scales. In order to further elaborate on this difference, several proposals to formalize the concept of macroscopic distinctness based on the “dead and alive” structure of a quantum state have been put forward [3–8]. For instance, the redundancy of information encoding in subparts of the system (like in the cells of the biological cat) [5], or the distance measured in units of “microscopic steps” [6] have been suggested. Even though these approaches are conceptually appealing, they suffer from some shortcomings. A general pure state does not have a Schrödinger-cat like structure, and, though one can always try to find a decomposition of a state into “dead and alive”, such a decomposition is never unique. Even in the case where a natural choice seems to exist, this may not automatically lead to the maximal result [9]. In addition, extensions to superposition with different weights, or to mixed states are not straightforward. This limits the proposals to analyze ideal situations, while experimental data is difficult to interpret.

Other measures are directly formulated for arbitrary quantum states [2, 10–14]. Some of them are based on a pre-chosen observable of the system and define general-

ized notions of “macroscopically distinct” as the spread of the wave function in the spectrum. The variance of this observable for pure states is closely connected to some proposals [10–13]. The more general approaches are however sometimes criticized to lack the conceptual beauty and clear physical intuition (as given by the distinctness of the two components for the other measures).

In this paper, we close the gap between these two basic approaches. We propose a measure that is applicable to superpositions of multiple states with unequal weights and is readily extendable to mixed states, thereby overcoming the shortcomings of previous proposals. We start with the intuition that $|\text{alive}\rangle$ and $|\text{dead}\rangle$ are macroscopically distinct if the two states can be distinguished by “classical” detectors [7], i.e. detectors that do not in general completely collapse the system into perfectly orthogonal states upon measurement but only weakly disturb the system. Needless to say that such detectors also do not perfectly extract information about any state of the system, hence they are said to have a limited resolution precision or resolution. Considering general pure states $|\Psi\rangle$ without specifying a subdivision into two branches, we quantify how much information about $|\Psi\rangle$ can be extracted by such detectors. The informative content is measured with the mutual information between system and measurement apparatus in the relevant bases. Then, we attribute an effective size to $|\Psi\rangle$ as the robustness of this information with respect to the detector’s resolution. The concept of macroscopic distinctness is hence formalized as “macroscopically extractable information”. This idea is generalized to mixed states via a convex roof construction (i.e., considering the “worst case”). In contrast to [8], which follows a similar approach using pairwise distinguishability, using the mutual information ensures a holistic treatment, and allows for a connection between different approaches.

The paper is summarized as follows. In Sec. II, we formalize the intuitive idea and define our measure for pure states. We discuss a paradigmatic example of multiple superposition states and establish a connection to the variance. In Sec. III, we present an extension of our measure to mixed states using a convex-roof construction, and illustrate it with a simple example. In Sec. IV we introduce alternative ways to formalize our intuition. For one of them, we use the basis-dependent discord. We

discuss implications and connections between the alternatives and to other proposals from the literature. In Sec. V, we provide a connection between macroscopic distinctness as formalized by our measure and fragility of entanglement to another system. We summarize and conclude in Sec. VI.

II. MACROSCOPICALLY EXTRACTABLE INFORMATION

A. Abstract definition

Consider a set $\mathbb{S} = \{\sqrt{p_\ell} |A_\ell\rangle\}_{\ell=1}^N$ of N quantum states $|A_\ell\rangle$ and N corresponding probabilities p_ℓ . This set can be used to define a superposition state

$$|\Psi\rangle_{\mathbb{S}} \propto \sum_{\ell=1}^N \sqrt{p_\ell} |A_\ell\rangle, \quad (1)$$

or of a “micro-macro” entangled state

$$|\Psi\rangle_{\text{mM}} = \sum_{\ell=1}^N \sqrt{p_\ell} |\ell\rangle_{\text{m}} |A_\ell\rangle_{\text{M}}, \quad (2)$$

between the system and some microscopic system with N orthogonal states $\{|\ell\rangle\}_{\ell=1}^N$ called “the atom” [15]. We wish to construct a meaningful definition for the size of such superposition states, based on some notion of generalized macroscopic distinctness of the superposed components \mathbb{S} .

Following [7], we assume that we measure the macroscopic system with a measurement device that has a rather coarse-grained resolution Δ (i.e., “low resolution” means large Δ). Let us consider a game in which Bob draws a random variable ℓ described by a probability distribution p_ℓ and sends the corresponding state $|A_\ell\rangle$ to Alice. She measures the received state with the detector (characterized by Δ), and obtains some outcome x . The information that she collects on the random variable held by Bob can be quantified by the mutual information (MI) of the probability distribution $p(\ell, x)$,

$$I_\Delta(A : \ell) = H(p_\ell) - \sum_x p(x) H(p(\ell|x)), \quad (3)$$

with the Shannon entropy $H(p_\ell) = -\sum_\ell p_\ell \log p_\ell$ [16]. Note that the MI can never exceed the Shannon entropy of the initial probability distribution $H(p_\ell)$. Hence, the maximal MI for a set of N orthogonal states with equal weights is given by $b_{\text{max}} = \log_2(N)$.

The intuition inherited from the macroscopic distinctness of the cat’s two states $|\text{dead}\rangle$ and $|\text{alive}\rangle$ tells us that a truly macroscopic superposition does not require technologically advanced detectors with high resolution in order to collapse the superposition to a single branch (or equivalently to learn the state of the atom in Eq. (2)) [7]. To quantify this intuition we define the effective size of $|\Psi\rangle_{\mathbb{S}}$ or $|\Psi\rangle_{\text{mM}}$ as the maximal Δ of the detector that still allows Alice to gain b bits of information about the preparation of Bob

$$\text{MIC}_b(\mathbb{S}) = \max_{\Delta} \{I_\Delta(A : \ell) \geq b\}, \quad (4)$$

standing for the *Macroscopicness of Information Content* of the superposition. The minimal information b is a parameter of the proposed measure, whose role we discuss later.

B. Model of a coarse-grained measurement

Up to this point we were quite unspecific about the measurement device. Indeed, the definition above only assumes that there is a meaningful way to attribute a resolving parameter Δ to the measurement device (and to continuously vary this parameter). In general, the detector does not have to be uniquely characterized by Δ , but can have additional knobs. In such a case, an additional optimization is necessary, as one is interested in the largest possible MI. However, we do not consider this more complicated situation in the following. As a first example, note that low resolution can come from inefficiencies modeled by a loss channel preceding an ideal measurement, in which case Δ is associated to the probability to not (or only partially) measure the system (see also [5, 17]).

In the following, however, we will consider the von Neumann pointer model with weak coupling between system and pointer. Suppose one would like to measure system with the observable

$$A = \sum_{\ell} a_{\ell} |A_{\ell}\rangle\langle A_{\ell}|, \quad (5)$$

which, for simplicity, is supposed to have non-degenerate discrete spectrum (if this is not the case replace the sum with an integral). For the formal definition of our measure the choice of A is irrelevant. However, it does determine which states are considered to be macroscopically distinguishable. Typically, we choose operators A with a classical limit such as collective spin operators for atomic ensembles or number of photons and quadrature operators for photonic state.

The measurement is done via a pointer P (i.e., an auxiliary system), which first interacts with the system and is subsequently read out in a preferred basis. Consider a pointer system modeled by a particle on a one-dimensional line with the usual commutation relation for position and momentum $[\hat{x}, \hat{p}] = i$ (with $\hbar = 1$). We assume the pointer’s initial state to be

$$|\xi_\Delta\rangle = \int \xi_\Delta(x) |x\rangle dx, \quad (6)$$

with Δ characterizing the width of the distribution $|\xi_\Delta(x)|^2$, and we choose a real valued function $\xi_\Delta(x)$. The system interacts with the pointer via the unitary $U = e^{-iA \otimes \hat{p}}$. Afterwards, the pointer is measured in the x -basis leaving the system in the state

$$\rho_x = \frac{K_x \rho K_x^\dagger}{\text{tr} K_x^\dagger K_x \rho} \quad (7)$$

with $K_x = \langle x|U|\xi_\Delta\rangle = \xi_\Delta(x - A)$. On an abstract level, this protocol realizes a general measurement with POVM elements $K_x^2 = \xi_\Delta^2(x - A)$. Trivially, if the width Δ of the initial pointer state tends to zero, one recovers the usual “strong” projective measurement $\xi_\Delta^2(x - A) \rightarrow \delta(x - A)$.

In contrast, the coupling becomes effectively weaker as Δ increases. The system is less disturbed by the measurement and, consequently, the measurement progressively loses resolution and becomes less informative. This is sometimes called a weak measurement.

In case one does not postselect on (or does not have access to) the measurement result x , the post-measurement state of the system –after tracing out the pointer– reads

$$\rho_{\text{out}} = \text{tr}_P U \rho \otimes |\xi_\Delta\rangle\langle\xi_\Delta| U^\dagger = \int \mu(p) e^{-ipA} \rho e^{ipA} dp, \quad (8)$$

where $\mu(p) = |\langle p|\xi_\Delta\rangle|^2$. In words, if the measurement outcome is ignored the effect of the weak measurement on the state is a dephasing channel generated by the observable A . Note that $\langle p|\xi\rangle$ and $\xi_\Delta(x)$ are connected via a Fourier transform, such that, in general, the weaker the measurement the lower is the strength of the induced dephasing.

To be more specific, we consider two examples for the pointer function $\xi_\Delta(x)$ in the following. In Sec. II C, we assume the distribution of the pointer to be square with a width Δ , such that an outcome x corresponds to a POVM element $E_\Delta(x - A) = \xi_\Delta^2(x - A)$ with

$$E_\Delta(x) = \begin{cases} \frac{1}{\Delta} & |x| \leq \frac{\Delta}{2} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Another important example is when $\xi_\Delta^2(x - A) = g_\Delta(x - A)$ is a Gaussian function with spread Δ , that is,

$$g_\Delta(x) = \frac{1}{\sqrt{2\pi}\Delta} e^{-\frac{x^2}{2\Delta^2}}. \quad (10)$$

C. Example: Equally spaced peaks

We now illustrate our formalism with a simple example of the equally weighted superposition

$$|\Psi\rangle_S = \frac{1}{\sqrt{k+1}} \sum_{\ell=0}^k \left| \ell \frac{N}{k} \right\rangle \quad (11)$$

of $k+1$ equally spaced eigenstates $A \left| \ell \frac{N}{k} \right\rangle = \frac{\ell N}{k} \left| \ell \frac{N}{k} \right\rangle$, all contained in the interval $[0, N]$, and a square pointed $E_\Delta(x)$ of Eq. (9). As the distribution is uniform one has $H(p_\ell) = \log_2(k+1)$. First, note that the probability to observe an outcome x only depends on k and the ratio $r = \frac{\Delta}{2N}$. So we directly move to the scale-invariant problem with $k+1$ eigenstates $\mathbb{S} = \{(1+k)^{-1/2} \left| \frac{\ell}{k} \right\rangle\}_{\ell=0}^k$ contained in the interval $[0, 1]$, and the square pointer $E_{2r}(y)$ of width $2r = \frac{\Delta}{N}$.

The calculation of the MI is mainly a combinatorial problem. Lengthy but straightforward arithmetics (see Appendix A) gives, for $r \geq 1/2$,

$$P_n^{\geq}(r, k) = \begin{cases} \frac{n}{k(k+1)r} & 1 \leq n \leq k \\ 1 - \frac{1}{2r} & n = k+1 \\ 0 & n > k+1 \end{cases} \quad (12)$$

and, for $r < 1/2$ with $c = \lfloor 2rk \rfloor$,

$$P_n^{<}(r, k) = \begin{cases} \frac{2n}{k(2(k+1)r)} & 1 \leq n \leq c-1 \\ \frac{c(k-c)(\frac{c}{k} + \frac{1}{k} - 2r) + \frac{2c}{k}}{2(k+1)r} & n = c \\ \frac{(c+1)(k-c+1)(2r - \frac{c}{k})}{2(k+1)r} & n = c+1 \\ 0 & n > c+1. \end{cases} \quad (13)$$

This implies

$$I_\Delta(A : \ell) = \begin{cases} \frac{N}{\Delta} (\log_2(k+1) - \frac{2 \log_2(H!(k))}{k(k+1)}) & \Delta \geq N \\ \sum_{n=1}^{c+1} P_n^{<}(\frac{\Delta}{2N}, k) \log_2(n) & \Delta < N \end{cases} \quad (14)$$

with the hyperfactorial $H!(k) = \prod_{n=1}^k n^n$. In Fig. 1 we plot I_Δ for several numbers of peaks, as well as the limiting case $k \rightarrow \infty$.

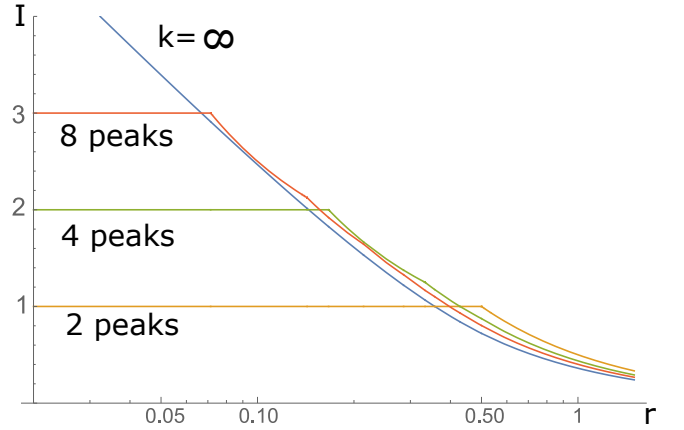


FIG. 1. Mutual information $I(A : \ell)$ as function of the ratio $r = \frac{\Delta}{2N}$ for two, four, eight and infinite number of peaks. While large MI can only be reached with many peaks and small r , the two-peak state maximizes the MI within this state family for $r \geq 1/2$.

For $\Delta \geq N$ the maximal MI is obtained for two peaks and is given by $I_\Delta = \frac{N}{\Delta}$. Accordingly for $b \leq 1$ this is also the state that maximizes

$$\max_k \text{MIC}_{b \leq 1} = \frac{N}{b} \quad (15)$$

For $\Delta < N$ things are more complicated, but numerical evidence shows that for $b = \log_2(k+1)$

$$\max_{k'} \text{MIC}_{b=\log_2(k+1)} = \frac{N}{2^b - 1} = \frac{N}{k} \quad (16)$$

is maximized by the state with $k+1$ peaks. Combining the two we find, for any b , the maximal size attained by state with equally spaced peaks in the interval $[0, N]$

$$\text{MIC}_{b,N}^{\text{peaks}} = \begin{cases} \frac{N}{b} & b \leq 1 \\ \frac{N}{2^b - 1} & b > 1 \end{cases} \quad (17)$$

To conclude this example let us remark that with the results above this family of states can be used for calibration of the measure MIC_b for any state. Concretely, for any superposition state in addition to attributing a value MIC_b for each b one says that the state under consideration is as macroscopic as $k+1$ equally spaced peaks in the interval $[0, N]$, for some k and N easily obtained from Eq. (17) and Eq. (16).

D. Role of b and calibration of the measure

The proposed measure is parametrized by b , that is, the amount of extractable information in the protocol of Sec. II A measured in bits. This might seem as a flaw of our approach, adding some arbitrariness to the definition. But this is not so, b can be understood as the “rank” of the macroscopic superposition – it counts the effective number of different components that are superposed. This is an important characterization of the state that is independent and irreducible to its “size”. For example, the state in the famous thought experiment of Schrödinger cat $|\uparrow\rangle|\text{alive}\rangle + |\downarrow\rangle|\text{dead}\rangle$ is undeniably a very large macroscopic superposition, still it is a superposition of only two components and can never yield more than one bit of information. Similarly, one can easily think of a microscopic state that is a superposition of many components yielding a large amount of information $b \gg 1$, nevertheless it has a small size $\text{MIC}_{b=1}$ even for one bit.

It is then appealing to introduce an archetypal reference state for each value of b , which can be used for the calibration of the size measure. In view of the results above, this can be naturally done using the family of $k+1$ -peaks states. Concretely, for any value of $b_k = \log_2(k+1)$ we can identify the state with $k+1$ equally spaced peaks in the interval $[0, N]$. Then for a general state $|\Psi\rangle_S$ and for each value b_k , in addition to attributing a value MIC_{b_k} , one can conclude that the state $|\Psi\rangle_S$ is as macroscopic as the state with $k+1$ peaks distributed on the interval of width

$$N_{b_k}(|\Psi\rangle_S) = \max_N \{N | \text{MIC}_{b_k, N}^{\text{peaks}} \leq \text{MIC}_{b_k} \}, \quad (18)$$

using the result of Eq. (17). $N_{b_k}(|\Psi\rangle_S)$ can be interpreted as a calibration of the size measure.

E. Connection to the variance

The variance of a state $V(|\Psi\rangle, A) = \langle \Psi | A^2 | \Psi \rangle - \langle \Psi | A | \Psi \rangle^2$ is a natural measure of how large is the spread of a state in the eigenbasis of A . So it is natural to study the relation of our measure to the variance. For this, we consider Gaussian pointers, Eq. (10), for which the MI can be expressed as

$$I_\Delta(A : \ell) = H(p(x)) - \sum_\ell p_\ell H(p(x|\ell)) \quad (19)$$

$$= H(p(x)) - \frac{1}{2} \log(2\pi e \Delta^2), \quad (20)$$

since $p(x|\ell) = g_\Delta(x - a_\ell)$. In Appendix D, we prove that MI is always upper bounded by the variance

$$I_\Delta(A : \ell) \leq \frac{V(|\Psi\rangle, A)}{(2 \ln 2) \Delta^2} \quad \forall \Delta. \quad (21)$$

One might wonder if there also exists a lower bound involving the variance. However, with the following example it is easy to see that no such bound can exist. For an appropriate choice of parameters p and N the superposition state $\sqrt{p}|0\rangle + \sqrt{1-p}|N\rangle$ can have an arbitrarily low MI and an arbitrarily large variance. Consequently, the two are inequivalent and the requirement for a large

MI is strictly more restrictive than for a large variance. Nevertheless, the inequality (21) becomes tight when Δ is sufficiently large

$$I_\Delta(A : \ell) \approx \frac{V(|\Psi\rangle, A)}{(2 \ln 2) \Delta^2}. \quad (22)$$

This shows that, for a weak Gaussian measurement and for pure states, our measure is connected to earlier proposals [10–13] where the variance $V(|\Psi\rangle, A)$ plays a role to measure the macroscopic distinctness. Equation (22) is further useful to evaluate our measure for small b .

III. MIXED STATES AND CONVEX ROOF

In practice, quantum states ρ are mixed. On the conceptual level, one can treat the *macroscopicness* and the *quantumness* of ρ as two independent aspects. The mixedness of a state ρ can then be attributed to the decay of its quantumness, while its macroscopicness, stemming from \mathbb{S} , is left unchanged. Nevertheless this is not satisfactory in our case. First, we would like the MIC measure to be a single quantity that encompasses both the macroscopicness and the quantumness of the state. Second, a mixed state $\rho = \sum q_i |\Psi_i\rangle\langle\Psi_i|$ admits infinitely many ensemble decompositions which can yield different average MIC, since different elements $|\Psi_i\rangle$ correspond to different \mathbb{S}_i and do not necessarily have the same size.

To get a MIC defined on all states ρ and non-increasing on average under mixing one uses the convex-roof extension

$$\widehat{\text{MIC}}_b(\rho) = \min_{\sum_k q_k |\Psi_k\rangle\langle\Psi_k| = \rho} \sum_k q_k \text{MIC}_b(|\Psi_k\rangle). \quad (23)$$

In words one finds the ensemble partition of $\sum q_i |\Psi_i\rangle\langle\Psi_i| = \rho$ that has the least average size, and defines this value as the size of ρ . This is by construction non-increasing under mixing, given any measure defined on pure states. Note that, despite the uncountable number of pure-state decompositions of ρ , the number of pure states in an extremal ensemble is limited to d^2 , where d is the rank of ρ . They also form a closed manifold, as there is a one to one mapping between decompositions of ρ and partitions of identity, or POVMs, see appendix B).

As an example, we consider quantum states lying in the span of two eigenstate of the observable $|0\rangle$ and $|N\rangle$ as in Sec. II C. The most general state of this form reads

$$\rho = \frac{1}{2} \begin{pmatrix} 1 + z_\rho & x_\rho - iy_\rho \\ x_\rho + iy_\rho & 1 - z_\rho \end{pmatrix}, \quad (24)$$

expressed in the basis $\{|0\rangle, |N\rangle\}$ in the superposition scenario, or $\{|0, 0\rangle_{\text{mM}}, |N, N\rangle_{\text{mM}}\}$ in the micro-macro entanglement scenario. To shorten the notation we define $\mathbf{r} = (x_\rho, y_\rho, z_\rho)$. As the size is invariant under rotation of the state around the z-axis, we assumed $y_\rho = 0$ in Eq. (24).

For pure states (i.e., $\mathbf{r}^2 = 1$) and a square pointer as defined in Eq. (9), the MI Eq. (3) can be easily computed

$$\tilde{I}_L(x) = \tilde{I}_0(x) \min\left(\frac{N}{\Delta}, 1\right) \quad (25)$$

with a convex function

$$\tilde{I}_0(x) = 1 - \log(|x|) + \frac{\sqrt{1-x^2}}{2} \log\left(\frac{1 - \sqrt{1-x^2}}{1 + \sqrt{1-x^2}}\right) \quad (26)$$

Consequently, for pure states the size is given by

$$\text{MIC}_b(x) = N \begin{cases} \frac{\tilde{I}_0(x)}{b} & I_0(x) > b \\ 0 & \text{otherwise,} \end{cases} \quad (27)$$

where the spread of the state N appears as a factor (as only the relative size of the spread to the pointer width is relevant).

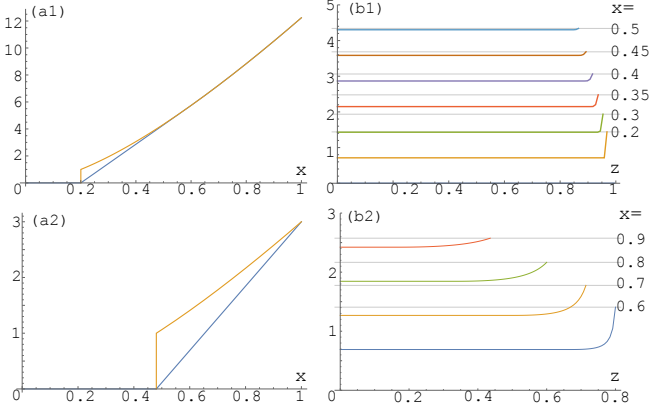


FIG. 2. (a1-a2) Rescaled size $\widehat{\text{mic}}_b(x_\rho, z_\rho)$ of the states as function of x_ρ for $b = 0.082$ (a1) and $b = 1/3$ (a2). The top curves corresponds to pure states with $z_\rho = \sqrt{1-x_\rho^2}$, and the bottom curve corresponds to states with $z_\rho = 0$. (b1-b2) Rescaled size $\widehat{\text{mic}}_b(x_\rho, z_\rho)$ of the state as function of z_ρ for $b = 0.082$ (b1) and $b = 1/3$ (a2). Different curves correspond to different constant values of x_ρ given on the right.

Given this expression one can compute the average size for any ensemble partition of ρ . It remains then to find the optimal ensemble. Though this step can be done analytically, it turns out to be quite tedious (see Appendix C). The final result is a function

$$\widehat{\text{MIC}}_b(x_\rho, z_\rho, N) = N \widehat{\text{mic}}_b(x_\rho, z_\rho), \quad (28)$$

that linearly depends on N , but has a complicated dependence of the state ρ .

In Figure 2 we plot the rescaled size $\widehat{\text{mic}}_b(x_\rho, z_\rho)$ for two values of $b = 0.082$ and $b = 1/3$. These value are chosen to allow a comparison with the example of [7], where equally-weighted pure superpositions $\frac{1}{\sqrt{2}}(|0\rangle + |N\rangle)$ are characterized by the probability P_c to correctly distinguish between the two branches ($|0\rangle \equiv |A\rangle$ and $|N\rangle \equiv |D\rangle$ in [7]; with the choice $P_c = 2/3$ in the example). The value of P_c does not uniquely determine the MI obtained by the measurement. Indeed, as entropy is concave given the two distributions $p(x|0)$ and $p(x|N)$, with the fixed guessing probability

$$P_c \equiv \frac{1}{2} \int \max(p(0|x), p(N|x)) dx = 2/3 \quad (29)$$

the MI can range from 0.082 bits for the extremal case, where for each outcome x either $p(0|x) = 2/3$ and $p(N|x) = 1/3$ or vice versa, to 1/3 bits for the other extreme, where for with probability 1/3 one obtains an outcome for which either $p(0|x) = 0$ or $p(N|x) = 0$ and $p(0|x) = p(N|x) = 1/2$ for all the others others.

IV. ALTERNATIVES AND COMPARISON

The convex-roof extension for mixed states, Eq. (4), is conceptually straightforward. However, the convex roof of an inverted function is generally difficult to handle in proofs or in calculations of specific examples. In this section, we present alternative formulations and compare them to recent contributions in the literature. We present two variants. For the first one, we start with the original idea using the MI, but do the convex-roof extension of the MI instead of MIC. The second alternative uses a slightly different motivation to directly measure the nonclassical part of the macroscopically extractable information. We find a formulation that turns out to be equivalent to the so-called basis-dependent quantum discord.

A. Direct convex-roof extension of the MI

Instead of the convex-roof extension of MIC we consider the direct convex roof of the MI

$$I_\Delta(A : \ell)_\rho = \min_{q_k, |\Psi_k\rangle} \sum_k q_k I_\Delta(A : \ell)_{|\Psi_k\rangle} \quad (30)$$

and define a slightly different version of the MIC, namely

$$\widehat{\text{MIC}}'_b(\rho) = \max\{\Delta | I_\Delta(A : \ell)_\rho \geq b\}. \quad (31)$$

For the example of Eq. (24) this definition gives the size of a pure state $\widehat{\text{MIC}}'_b(\rho) = \text{MIC}_b(x_\rho)$ with the same x_ρ in Eq. (27), as it follows from the convexity of $\tilde{I}_0(x)$. So it has the advantage to be more straightforward to compute. In addition this alternative definition allows us find the following connection.

In [18], a set of criteria were proposed for quantities that aim to capture the macroscopic coherence of a state. These are in the same spirit as the criteria for good entanglement measures. The most important ones say that a valid measure should (C1) vanish if and only if a state is an “incoherent” mixture of the form $\sum_\ell p_\ell |A_\ell\rangle\langle A_\ell|$; (C2) not increase under any “covariant” operation. An operation is covariant when it commutes with transformations of the form e^{-itA} – this set captures all the possible operations which cannot create a superposition of the $|A_\ell\rangle$ and which respect the “scale” $|a_i - a_j|$ of a superposition $|A_i\rangle + |A_j\rangle$. (C2) can be broken down into two versions, (C2a) for deterministic processes, and (C2b) for stochastic processes under which the measure cannot increase on average. In addition, one can demand that a measure be (C3) convex (i.e., non-increasing under mixing) and (C4) increasing with respect to the scale $|a_i - a_j|$.

We show in Appendix E that this extended $I_\Delta(A : \ell)_\rho$, assuming a Gaussian pointer, satisfies all criteria (C1-4). In addition, $\widehat{\text{MIC}}'_b$ satisfies (C2a), (C4) and a modified version of (C1), namely

$$\widehat{\text{MIC}}'_b = 0 \Leftrightarrow I_\Delta(A : \ell) \leq b \forall \Delta. \quad (32)$$

In other words, $\widehat{\text{MIC}}$ is well-behaved in the sense that it vanishes for states that are close to incoherent mixtures, cannot increase under covariant operations, and is increasing with the scale of a superposition.

It is also worth noting that the relations between the MI and the variance for pure states of Eq. 21 are directly generalized to mixed states via the convex roof of MI

$$I_\Delta(A : \ell)_\rho \leq \frac{\mathcal{F}(\rho, A)}{(8 \ln 2) \Delta^2} \quad (33)$$

$$I_\Delta(A : \ell)_\rho \approx \frac{\mathcal{F}(\rho, A)}{(8 \ln 2) \Delta^2} \quad \text{for } b \rightarrow 0. \quad (34)$$

Where the quantum Fisher information $\mathcal{F}(\rho, A)$ [19] of the state ρ with respect to the operator A is known to equal to four times the convex roof of the variance [20]. It follows that $\widehat{\text{MIC}}'_b$ satisfies

$$\widehat{\text{MIC}}'_b(\rho) \leq \sqrt{\frac{\mathcal{F}(\rho, A)}{(8 \ln 2) b}} \quad \text{and} \quad (35)$$

$$\widehat{\text{MIC}}'_b(\rho) \approx \sqrt{\frac{\mathcal{F}(\rho, A)}{(8 \ln 2) b}} \quad \text{for } b \rightarrow 0. \quad (36)$$

This allows one to obtain upper-bounds on the size. Moreover, since the quantum Fisher information is known to satisfy all criteria (C1-4), we conclude that $\widehat{\text{MIC}}'_b$ fulfills them in the limit of $b \rightarrow 0$. Note that the quantum Fisher information plays a central role in one of the general proposals for macroscopic distinctness [13], so this measure is in some sense contained in the presented family as a limiting case.

In particular, the insights we have about the quantum Fisher information can be used to apply our measure for small b to real experimental data [21].

B. Alternative measure using quantum correlations

In this section, we build up an alternative measure which is conceptually similar but has a slightly different motivation. As argued earlier, the distinguishability of a set of states under a noisy measurement can be captured by the mutual information between Bob, who prepares the ensemble of states, and Alice, who reads out the measurement device. Put differently, when a measurement device interacts with a system in the superposition state $|\Psi\rangle_S = \sum_\ell \sqrt{p_\ell} |A_\ell\rangle$, the correlations $I(A : \ell)$ between the macroscopic system M and measurement device P are related to how well the device discriminates the branches of the superposition. However, $I(A : \ell)$ can be non-zero even when the system is initially in an incoherent mixture $\sum_\ell p_\ell |A_\ell\rangle\langle A_\ell|$. This issue can be avoided by using the convex roof constructions for mixed states. This is conceptually appealing, but comes at a price of make things hard to compute, as illustrated with the example of Section II C. But is there a more direct way to avoid this issue that retains the nice physical intuition behind the MI.

Here we introduce the quantity $C_\Delta(\rho, A)$ that is related to $I_\Delta(A : \ell)$ in spirit, but can be directly applied to mixed states. We start by introducing it in two different ways, and then show that they are equivalent. Let the system start in an arbitrary state ρ , and the measurement device in the initial pointer state $|\xi_\Delta\rangle$. We call

$$\rho' = U \rho \otimes |\xi_\Delta\rangle\langle \xi_\Delta| U^\dagger \quad (37)$$

the overall state after the interaction $U = e^{-iA \otimes \hat{p}}$, ρ'_M the post-measurement state of the system, given in Eq. (8), and ρ'_P the final state for the pointer. Using the Von Neumann entropy $S(\rho) = -\text{Tr} \rho \log \rho$, define

$$C_\Delta(\rho, A) := S(\rho'_M) - S(\rho) \quad (38)$$

as the entropy difference between the post-measurement state ρ'_M , given in Eq. (8), and the initial state ρ . Intuitively, the entropy increase in the system can only come from its correlations to the pointer created by the interaction. Hence, $C_\Delta(\rho, A)$ captures how much information is potentially available to the pointer about the system [22]. Note that this quantity avoids all problems associated with mixed states. In particular, the system state with no coherence

$$\mathcal{G}(\rho) := \sum_\ell |A_\ell\rangle\langle A_\ell| \rho |A_\ell\rangle\langle A_\ell| = \sum_\ell p_\ell |A_\ell\rangle\langle A_\ell| \quad (39)$$

is not affected by the interaction with pointer, implying $C_\Delta(\mathcal{G}(\rho), A) = 0$.

An alternative definition can be given via the quantum mutual information (QMI) $\mathcal{I}(P : M)_{\rho'} := S(\rho'_P) + S(\rho'_M) - S(\rho')$ between the system and the pointer after the interaction. As we show in the next paragraph,

$$C_\Delta(\rho, A) := \mathcal{I}(P : M)_{\rho'} - \mathcal{I}(P : M)_{\mathcal{G}(\rho)}. \quad (40)$$

is given by the QMI for the initial state ρ minus the QMI for its incoherent version $\mathcal{G}(\rho)$. Here, the issue of mixed states is resolved even more explicitly as the incoherent contribution to the QMI is simply subtracted. In fact, the definition (40) also makes it clear that $C_\Delta(\rho, A)$ corresponds to the (fixed-basis) quantum discord of the final state [23]. As the quantity

$$\mathcal{I}(P : M)_{\mathcal{G}(\rho')} = S(P|M) = \sum_\ell p_\ell S(\rho_{P|\ell}), \quad (41)$$

with $\rho_{P|\ell} = \frac{\text{tr}_M \rho' |A_\ell\rangle\langle A_\ell|}{\text{tr} \rho' |A_\ell\rangle\langle A_\ell|}$ and $\text{tr} \rho' |A_\ell\rangle\langle A_\ell| = p_\ell$, is equal to the conditional entropy of the pointer on the system measured in the eigenstates of A (note that the measurement commutes with the interaction). So since the initial state of the pointer is pure $\mathcal{I}(P : M)_{\mathcal{G}(\rho')} = \mathcal{J}(P|M)_{\rho'}$ gives the QMI between the system and the pointer, available upon the measurement of A . Note that the usual approach is to maximize the classical correlations over all possible measurements on M ; but since we have a fixed observable A of interest here, it is natural to fix the measurement.

Let us now show that the two definitions are equivalent. Since the interaction is unitary, we have $S(\rho') = S(\rho \otimes |\xi_\Delta\rangle\langle \xi_\Delta|) = S(\rho)$, thus $\mathcal{I}(P : M)_{\rho'} = S(\rho'_M) + S(\rho'_P) - S(\rho')$. For the classical correlations, we write $\sigma = \mathcal{G}(\rho)$ and $\sigma' = \sum_\ell p_\ell |A_\ell\rangle\langle A_\ell| \otimes e^{-ia_\ell p} |\xi_\Delta\rangle\langle \xi_\Delta| e^{ia_\ell p}$. It follows that

$$\begin{aligned} \mathcal{J}(P|M)_{\rho'} &= S(\sigma'_M) + S(\sigma'_P) - S(\sigma') \\ &= H(p_\ell) + S(\rho'_P) - H(p_\ell) \\ &= S(\rho'_P). \end{aligned} \quad (42)$$

Therefore $C_\Delta(\rho, A) = S(\rho'_M) + S(\rho'_P) - S(\rho) - S(\rho'_P) = S(\rho'_M) - S(\rho)$.

So now we have the quantity $C_\Delta(\rho, A)$ in place of $I_\Delta(A : \ell)$. Note that, in the case of pure states, there is an inequality

$$C_\Delta(|\Psi\rangle, A) \geq I_\Delta(A : \ell). \quad (43)$$

This is because a pure post-interaction state of M and P has $\mathcal{I}(P : M)_{\rho'} = 2\mathcal{J}(P|M)_{\rho'} = 2S(\rho'_M)$, so $C_\Delta(|\Psi\rangle, A) = J(P|M)$. Then observe that $I_\Delta(A : \ell)$ measures the correlations of σ' with respect to a measurement of the pointer observable – which cannot exceed the mutual information $\mathcal{I}(P : M)_{\mathcal{G}(\rho')}$.

The rest of this section is devoted to examining the properties of $C_\Delta(\rho, A)$. We assume a Gaussian pointer from now on. As before, we can define another version of MIC:

$$\widetilde{\text{MIC}}_b(\rho) = \max\{\Delta | C_\Delta(\rho, A) \geq b\}. \quad (44)$$

Just as before, we find that C_Δ satisfies all the coherence measure criteria (C1-4) – see Appendix G for the proof. Again, $\widetilde{\text{MIC}}_b$ satisfies (C2a), (C4) and a modified version of (C1),

$$\widetilde{\text{MIC}}_b(\rho) = 0 \Leftrightarrow C_\Delta(\rho, A) \leq b \forall \Delta. \quad (45)$$

Let us look at the behaviour of C_Δ in the limit of a weak Gaussian measurement, where Δ is large. We find that, for pure states, to leading order,

$$C_\Delta(|\Psi\rangle, A) \approx \frac{h(\Delta^{-2})}{4} V(|\Psi\rangle, A), \quad (46)$$

where $h(t) = -t \log t$. This contrasts a little with Eq. (22) for $I_\Delta(A : \ell)$, where the leading order was Δ^{-2} . See Appendix H for the proof.

Recently, another measure for macroscopic distinctness has been proposed [24]. Although formulated differently, it is closely related to the quantity $C_\Delta(\rho, A)$. In particular, both measures fulfill the proposed set of criteria for macroscopic coherence [18]. As noted in [24], small Δ leads to the counter-intuitive result that the measure assigns a larger value to some product states than to a superposition of two extremal eigenstates of A . This reveals the role of Δ in these measures as a characteristic scale. Using the measure tells us how much quantum coherence the state of interest provides on this scale. In our framework it is $\widetilde{\text{MIC}}_b(\rho)$, rather than $C_\Delta(\rho, A)$, that is quantifying the macroscopic quantumness of the state. But along the same line $\widetilde{\text{MIC}}_b(\rho)$ can be understood as the maximal scale at which the state provides the desired amount of coherence.

V. IMPLICATIONS ON FRAGILITY

In this section we consider the micro-macro entangled state

$$|\Psi\rangle_{\text{mM}} = \sum_{\ell=1}^N \sqrt{p_\ell} |\ell\rangle_{\text{m}} |A_\ell\rangle_{\text{M}}.$$

Under the assumption $\langle A_j | A_k \rangle = \delta_{jk}$ Eq. (2) gives the Schmidt decomposition of $|\Psi\rangle_{\text{mM}}$. For the particular

case $\frac{1}{\sqrt{2}}(|0\rangle_{\text{m}} |A\rangle_{\text{M}} + |1\rangle_{\text{m}} |D\rangle_{\text{M}})$ we know that the micro-macro entanglement is more fragile for a larger size of the macroscopic part of the state using the framework of [7, 17]. Here we will show that a similar relation between the size of the state as defined in Eq. (4) and the fragility of entanglement under certain type of noise persists in the general case. The intuition behind is rather simple: If the noise channel can be interpreted as an imprecise measurement of the system by the environment, then the size of the state relates to the amount of information extractable by the environment. The decay of entanglement through the channel is related to the information obtained by the environment. Since at least the mathematical modeling of the environment and a measurement pointer is similar, we denote the environment as P as well.

A. Entanglement of formation

Concretely, we consider the entanglement of formation $E_F(\rho_{AB})$. E_F is an entanglement measure [25] on bipartite states, defined as the convex roof of the entropy of entanglement

$$E_F(\rho_{AB}) = \min_{\sqrt{q_k} |\Psi_k\rangle} \sum_k q_k S(\rho_k^B), \quad (47)$$

where the entropy of entanglement is by definition an entanglement measure on bipartite pure states given by the Von Neumann entropy $S(\rho_k^B) = S(\rho_k^A)$ of the partial states $\rho_k^B = \text{tr}_A |\Psi_k\rangle \langle \Psi_k|$. Because we assume all the branches of $|\Psi\rangle_{\text{mM}}$ to be orthogonal, its entanglement of formation reads

$$E_F(|\Psi\rangle_{\text{mM}}) = H(p_\ell). \quad (48)$$

Note that the entanglement in the state is invariant under local unitary transformation, so its amount is independent of the spread of the state in the spectrum of A and of the macroscopicness of the state.

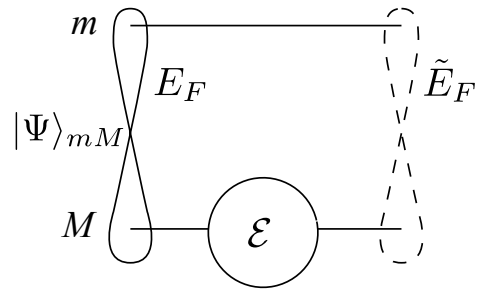


FIG. 3. Noise channel \mathcal{E}_X^A acting on the macroscopic part of the state $|\Psi\rangle_{\text{mM}}$.

B. Noise as measurement by environment

For any Kraus representation \mathbf{K} of a channel \mathcal{E}

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_x K_x \rho K_x^\dagger \\ &= \text{tr} \left(\sum_x K_x \otimes |x\rangle \langle 0|_P \rho \otimes |0\rangle \langle 0|_P \left(\sum_x K_x^\dagger \otimes |0\rangle \langle x|_P \right) \right) \end{aligned} \quad (49)$$

can be interpreted as a measurement of the system by the environment, described by the POVM elements induced by the Kraus operators $\{E_x = K_x^\dagger K_x\}$ (in the expression above the sum is replaced by an integral if a Kraus representation is continuous). For simplicity, the channel is supposed to act only on the macroscopic part (see Fig. 3). Hence, the output of the channel

$$\mathcal{E}(\rho) = \sum_x p(x) \rho_x \quad (50)$$

is a mixture of states $\rho_x = \frac{K_x \rho K_x^\dagger}{p(x)}$ with $p(x) = \text{tr} E_x \rho$ conditional on the environment observing outcome x . Note that the POVM elements do not uniquely specify the channel, as the same element E_x can correspond to physically different Kraus operators $U K_x = \sqrt{E_x}$.

Now consider the action of the channel on the state $\rho = |\Psi\rangle\langle\Psi|_{\text{mM}}$, for which all the conditional states

$$\rho_x = |\Psi_x\rangle\langle\Psi_x| \quad \text{with} \quad |\Psi_x\rangle = \frac{K_x |\Psi\rangle_{\text{mM}}}{\sqrt{p(x)}} \quad (51)$$

are also pure. Similarly to Eq. (3) we define the MI between the microscopic system and the measurement (that arises from the Kraus representation \mathbf{K} of the channel) carried by the environment $I_{\mathcal{E},\mathbf{K}}(P : \ell)$. One has

$$\int p(x) H(p(\ell|x)) dx = H(p_\ell) - I_{\mathcal{E},\mathbf{K}}(P : \ell), \quad (52)$$

where $p(\ell|x) = \text{tr} |\ell\rangle\langle\ell|_m \rho_x$. Note that it does not matter whether the projection $\{|\ell\rangle\langle\ell|\}$ on the atom's side or the measurement $\{E_x\}$ by the environment is performed first. The Shannon entropy of the distribution $p(\ell|x)$ upper-bounds the Von Neumann entropy of the partial states $\rho_x^M = \text{tr}_m |\Psi_x\rangle\langle\Psi_x|$ and $\rho_x^m = \text{tr}_M |\Psi_x\rangle\langle\Psi_x|$ [26]

$$S(\rho_x^M) = S(\rho_x^m) \leq H(p(\ell|x)). \quad (53)$$

This inequality allows one to obtain a bound on the average partial entropy of the post-channel state

$$\sum_x p(x) S(\rho_x^M) dx \leq H(p_\ell) - I_{\mathcal{E},\mathbf{K}}(P : \ell). \quad (54)$$

The left hand side is the average entropy of entanglement of the state $\mathcal{E}(|\Psi\rangle_{\text{mM}})$ that correspond to its pure-state partition provided by the Kraus representation \mathbf{K} . Consequently, by definition of the entanglement of formation one has

$$E_F(\mathcal{E}(|\Psi\rangle_{\text{mM}})) \leq E_F(|\Psi\rangle_{\text{mM}}) - I_{\mathcal{E},\mathbf{K}}(E : \ell). \quad (55)$$

In words, the decay of entanglement of formation through a channel is lower or equal than the MI obtained by the environment via the measurement induced by *any* Kraus representation by the channel.

C. Examples

(i) Dephasing generated by the observable A

$$\mathcal{E}_X^\delta(\rho) = \int \mu(\lambda) e^{-i\lambda A} \rho e^{i\lambda A} d\lambda, \quad (56)$$

of strength δ given by the width of the distribution $\mu(\lambda)$. As already mentioned in (8), this noise corresponds to a coarse-grained measurement of A by the environment. And the noise distribution $\mu(p) = |\langle p|\xi\rangle|^2$ is related to the resolution of the measurement $\xi_\Delta^2(x) = |\langle x|\xi\rangle|^2$ by a Fourier transform implying $\delta \sim \frac{1}{\Delta}$. This shows that the quantity $I_\Delta(A : \ell)$ yields a lower bound on the decrease of entanglement in the state after the action of the channel \mathcal{E}_X^Δ . Similarly, $1/\text{MIC}_b(|\Psi\rangle_{\text{mM}})$ gives an upper bound on the amount of noise $\delta \sim \frac{1}{\Delta}$ that leaves $E_F(|\Psi\rangle_{\text{mM}}) - b$ bits of entanglement in the system.

In the case of a channel \mathcal{E} describing weak Gaussian noise from the environment, Eq. (22) shows that

$$E_F(\mathcal{E}(|\Psi\rangle_{\text{mM}})) \lesssim E_F(|\Psi\rangle_{\text{mM}}) - \frac{V(|\Psi\rangle_{\text{mM}}, A)}{(2 \ln 2) \Delta^2}. \quad (57)$$

It also turns out that C_Δ lets us say something about the degradation of quantum correlations between m and M . Note that $\mathcal{E}(|\Psi\rangle_{\text{mM}})$ has the structure of a “maximally correlated state”, which is generally written as $\sum_{i,j} \rho_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j|$. It is known that the entanglement of a maximally correlated state is often the same as the coherence of the corresponding single-system state $\sum_{i,j} \rho_{ij} |i\rangle\langle j|$ [27–29]. For example, this is true for the distillable entanglement E_D [25] and the relative entropy of coherence C_R [30], which can be written as

$$C_R(\rho) = S(\rho || \mathcal{G}(\rho)), \quad (58)$$

that is, the relative entropy between a state and its fully dephased version. It also has the simple expression $C_R(\rho) = S(\mathcal{G}(\rho)) - S(\rho)$.

The channel \mathcal{E} leaves the fully dephased part of a state unchanged. Therefore we simply have

$$\begin{aligned} E_D(\mathcal{E}(|\Psi\rangle_{\text{mM}})) &= C_R(\mathcal{E}(|\Psi\rangle_{\text{mM}})) \\ &= C_R(|\Psi\rangle_{\text{mM}}) - C_\Delta(|\Psi\rangle_{\text{mM}}, A) \\ &= E_D(|\Psi\rangle_{\text{mM}}) - C_\Delta(|\Psi\rangle_{\text{mM}}, A) \\ &\approx E_D(|\Psi\rangle_{\text{mM}}) - \frac{h(\Delta^{-2})}{4} V(|\Psi\rangle_{\text{mM}}, A), \end{aligned} \quad (59)$$

where the final line uses the approximation (46) for a weak measurement.

(ii) A loss channel \mathcal{L}_η with “efficiency” η (corresponding to the efficiency of the measurement device) models a process where each particle (subsystem) is lost to the environment with probability $1 - \eta$, in other words the initial states of the particle and environment are eventually swapped. This symmetry implies that the transmitted state of the system $\mathcal{L}_\eta(\rho)$ and the state of the partial state of environment $\rho_{\eta'}^E$ are the same if $\eta' = 1 - \eta$ (the roles of the “transmitted” and “reflected” systems are exchanged). For a family of states \mathbb{S} one defines $I_\eta(A : \ell)$ as the maximal MI that is obtainable with a measurement device of efficiency η , and the corresponding $\text{MIC}_b(\mathbb{S})$ as the minimal efficiency that allows to obtain b bits. Again, for the state $|\Psi\rangle_{\text{mM}}$ the quantity $I_{1-\eta}(A : \ell)$ gives a bound on the decay of entanglement through the loss channel \mathcal{L}_η , while $1 - \text{MIC}_b$ is the minimal transmission of the channel that leaves at least $E_F(|\Psi\rangle_{\text{mM}}) - b$ bits of entanglement in the system.

VI. CONCLUSION AND DISCUSSION

Starting with the intuition that the macroscopic distinctness between two states “dead and alive” can be understood as “the ease to distinguish” the two states we lift this intuition to superpositions of multiple components by looking at “the ease to obtain information” about the state. We formalize this idea into a general measure that is also useful for mixed states. More precisely, we first quantify how much information one can extract from a pure state by measuring it with a classical detector with a limited resolution. Second, the minimal resolution that is required to extract the desired amount of information b is associated with a measure that quantifies the “macroscopic distinctness” of the state, that we call macroscopicity of information content (MIC). Throughout a large part of the paper we use the Von Neumann model for a weak measurement of a fixed observable A to model the classical detector.

To extend our measure to mixed states we use a convex roof construction, and illustrated it on a simple example. It is argued that the parameter b in our family of measures attributes a kind of “macroscopicity rank” to the superposition as it counts the effective number of components that are superposed. We also establish a relation between our measure and the variance of the state with respect to the operator A (its quantum Fisher information for a mixed state), that plays a central role in previously defined measures. In particular, we show that for a Gaussian pointer they are equal in the limit of small b .

Later, we present an alternative formulation of our measure, which stems from the same intuition but allows to directly deal with mixed states without the detour of a heavy convex roof construction. It turns out to be equal to the basis-dependent quantum discord, and is also closely related to the measure for macroscopic distinctness that has been proposed in [24]. In particular, it also fulfills the proposed set of criteria for macroscopic coherence [18]. So we can interpret it as the maximal scale at which the state provides the required amount of coherence, as quantified by b .

Finally, we study the relation between the fragility of the state and its macroscopicity quantified by our measures. Concretely, we analyze the decay of the entanglement of formation in a micro-macro state when noise is applied on the macro side. We show that regardless of the model of the classical detector used to quantify the size, there is always a noise channel for which the fragility of entanglement is directly related to the macroscopicity of the state. This result is then applied to two models of classical detector: a weak measurement of A central to the paper, and a generic inefficient detector modeled by a loss channel preceding an unknown measurement.

Our work provides a novel tool to analyze and compare recent and future experiments aiming at the observation of quantum effects at larger and larger scales.

Acknowledgments.— We thank Nicolas Sangouard for interesting discussions. This work was supported by the National Swiss Science Foundation (SNSF) projects P2GEP2.151964 and 00021.149109, and the Austrian Science Fund (FWF) P28000-N27, and the European Research Council (ERC MEC).

Appendix A: Details to equally-spaced-peaks example, Sec. II C

The following calculation gives some details for the example discussed in Sec. II C. The probability of a measurement outcome y is given

$$p(y) = \frac{\#_{\epsilon}^{(a,b]}}{2r(k+1)}, \quad (\text{A1})$$

where $\#_{\epsilon}^{(a,b]}$ counts the number of peaks (elements of \mathbb{S}) in the interval $(a, b]$. Its value can be expressed as the difference $\#_{\epsilon}^{(a,b]} = \#_{\leq}^b - \#_{\leq}^a$ of the number of peaks with $y \leq b$ (respectively $y \leq a$), which reads

$$\#_{\leq}^y = \begin{cases} 0 & y < 0 \\ \lfloor yk \rfloor + 1 & 0 \leq y < 1 \\ k + 1 & 1 \leq y \end{cases} \quad (\text{A2})$$

with $\lfloor x \rfloor$ denoting the integer part of x . The outcome y can be equally well triggered by any peak from the interval $(y - r, y + r]$, hence the conditional entropy is given by

$$H(p(\ell|y)) = \log_2 \left(\#_{\epsilon}^{(y-r, y+r]} \right). \quad (\text{A3})$$

The MI reads

$$I(A : \ell) = \log_2(k+1) - \int p(y) \log_2 \left(\#_{\epsilon}^{(y-r, y+r]} \right) dy \quad (\text{A4})$$

Since both the probability of an outcome and the conditional entropy are uniquely determined by the number of peaks in the corresponding interval, one can rephrase the problem in terms of the random variable n that englobes all the outcomes compatible with n peaks. One has

$$P_n = \int \delta_{\#_{\epsilon}^{(y-r, y+r]} = n} p(y) dy \quad (\text{A5})$$

and

$$I(A : \ell) = \log_2(k+1) = \sum_{n=0}^{n_{\max}} P_n \log_2(n). \quad (\text{A6})$$

Appendix B: Finite number of pure states in extremal ensemble

Here, we discuss a simplification in the convex-roof construction for measures defined for mixed states. Let us assume that ρ is a full-rank state. If this is not true, one simply restricts the Hilbert space to the support of ρ . The number of ensemble averages of any non-pure density matrix ρ is infinite. Moreover, even the number of pure states in such an ensemble is not bounded, one can even have ensembles defined by a non-discrete probability density on the manifold of pure states. This being said, the number of pure state in an *extremal* ensemble is actually limited to d^2 , where d is the rank of ρ . This can be seen from the following argument.

First, there is a one-to-one correspondence between decompositions of ρ (not necessarily in pure states) and

partitions of identity (or POVMs) known as ρ -distortion [31]. For each POVM, $\{E_i\}$ with $\sum_i E_i = \mathbb{1}$, the operator $\rho_i = \frac{\sqrt{\rho} E_i \sqrt{\rho}}{\text{tr} \rho E_i}$ is a valid state, and

$$\sum_i \underbrace{p_i}_{\text{tr} \rho E_i} \rho_i = \sqrt{\rho} \sum_i E_i \sqrt{\rho} = \rho. \quad (\text{B1})$$

Second, ρ_i is a pure state iff E_i is rank one. This yields a one-to-one correspondence between ensemble partition (pure states) of ρ and POVM composed of rank-one operators $\{E_i\}$. Finally, it is known that in dimension d an extremal POVM, i.e. that a measurement that does not correspond to a mixture of different POVMs (such a procedure physically corresponds to randomly choosing the measurement to perform and forgetting the choice), has maximally $n = d^2$ elements [32]. Via the correspondence above the same holds for extremal ensemble decomposition of ρ , and by construction the minimal size of Eq. (23) is attained by an extremal ensemble.

Appendix C: Convex roof example of Sec. III

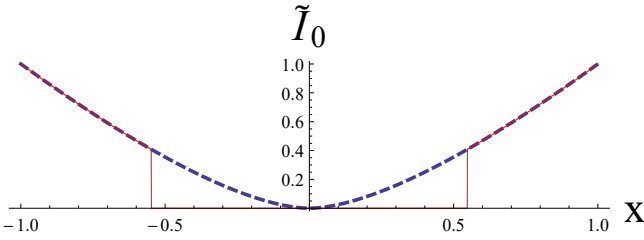


FIG. 4. Mutual information for pure states $I_0(x)$ (dashed, thick) and a rescaled size $\frac{b}{N} \text{MIC}_b$ for $b = 0.4$ (solid thin) as functions of x .

a. Restriction to the XZ plane Consider an ensemble decomposition $\sum_i q_i |\Psi_i\rangle\langle\Psi_i| = \sum_i \frac{q_i}{2} (\mathbb{1} + \mathbf{r}_i \cdot \boldsymbol{\sigma}) = \rho$. We show that there exists another decomposition that has a smaller or equal size but only involves states that lie in the XZ plane. To do so notice that

$$\sum q_i \bar{\rho}_i = \sum_i \frac{q_i}{2} (\mathbb{1} + \bar{\mathbf{r}}_i \cdot \boldsymbol{\sigma}) = \rho \quad (\text{C1})$$

also holds for each $\bar{\mathbf{r}}_i = (x_i, 0, z_i)$ restricted to the XZ plane. This is not a partition in pure state, but it naturally gives one, since each $\bar{\rho}_i$ can be decomposed in pure state as $\bar{\rho}_i = \lambda_i |\Psi_i^{(+)}\rangle\langle\Psi_i^{(+)}| + (1 - \lambda_i) |\Psi_i^{(-)}\rangle\langle\Psi_i^{(-)}|$ with the corresponding Bloch sphere vectors $\mathbf{r}_i^{(\pm)} = (x_i, 0, \pm \sqrt{1 - x_i^2})$. Finally, one has

$$\text{MIC}_b(|\Psi_i\rangle) \geq \lambda_i \text{MIC}_b(|\Psi_i^{(+)}\rangle) + (1 - \lambda_i) \text{MIC}_b(|\Psi_i^{(-)}\rangle) \quad (\text{C2})$$

since $|\sqrt{x_i^2 + y_i^2}| \geq |x_i|$ and the size is monotonously increasing. Consequently the new decomposition

$$\rho = \sum_i q_i \sum_{s=\pm 1} (1 - (-1)^s (1 - \lambda_i)) |\Psi_i^{(\text{sign}(s))}\rangle\langle\Psi_i^{(\text{sign}(s))}| \quad (\text{C3})$$

yields a lower or equal size. It follows than from the beginning it is sufficient to only consider the ensembles

where all elements lie in the XZ plane. As follows from [], extremal ensembles of this form involve three state at most.

b. Optimal ensemble Recall that the size of pure states in Eq. (27) is zero for small $|x|$ (such that $\tilde{I}_0(x) < b$) and then increases monotonously with $-|x|$. In addition, $I_0(x)$ is convex in the regions $(r \equiv I_0^{-1}(b), 1]$ and $[-1, -r)$.

If $|x_\rho| \leq r$, then the size of the state is zero. For example, this follows from the *vertical* decomposition: ρ is a mixture of the two pure states that have the same $x = x_\rho$, that both have zero size. So in the following we assume $x_\rho > r$. Without loss of generality, it follows that in the ensemble decomposition of ρ there is at least one pure state that lies in the right white sector of the XZ circle on the right from ρ , see Fig. 5. Actually, there are two possibilities: either (i) all the states lie in the right white sector, i.e. all these states satisfy $x \in (r, 1]$, or (ii) some lie outside.

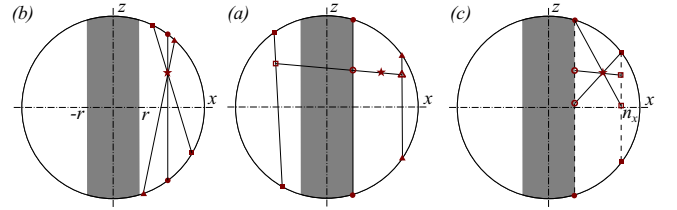


FIG. 5. A cut of the Bloch sphere through the XZ plane. The star represents the state ρ . (a) The convexity argument implies that vertical decomposition (circles) yields the lowest size among the three decomposition represented by line crossing ρ . (b) The ensemble consisting of states represented by triangles and circles always yields a lower size than the ensemble depicted by triangles and squares. (c) All the 3 ensembles give the same size.

The case (i) is rather simple, the convexity of the size in the $(r, 1]$ region implies that the ensemble with the minimal average size should have all the pure states with the same x as ρ , again this is fulfilled by the vertical decomposition, see Fig. 5(a), and gives $\text{MIC}_b^{\text{vert.}}(\rho) = \text{MIC}_b(x_\rho)$.

The case (ii) is more involved, however it can be simplified by the following remark. Let us label all the pure in the ensemble states with $x \leq r$ by $|\Phi_i\rangle$ and all states the states with $x > r$ by $|\Psi_i\rangle$. We then have

$$\rho = \underbrace{\sum_i q_i |\Psi_i\rangle\langle\Psi_i|}_{\equiv p \sigma} + \underbrace{\sum_i \bar{q}_i |\Phi_i\rangle\langle\Phi_i|}_{\equiv (1-p) \tau}, \quad (\text{C4})$$

with both σ and τ valid density matrices, that are represented by the triangle and the empty square in Fig. 5(b). The figure also directly suggests a decomposition that has a smaller average size that the one we started with. This is given by

$$\rho = p' \sigma + (1 - p') \tau', \quad (\text{C5})$$

where τ' (represented by the empty circle in Fig. 5(b)) lies on the intersection of the $x = r$ line and the line passing through τ , ρ and σ and has a zero average size (think of its vertical decomposition). In addition, one easily sees that $p' \leq p$ implying that this decomposition indeed yields a lower average size.

The two previous observation imply that minimal ensemble consists of at most of four pure state: the two states $|\Psi_r^{(+)}\rangle$ and $|\Psi_r^{(-)}\rangle$ with $x = r$ and the two states $|\Psi_{n_x}^{(+)}\rangle$ and $|\Psi_{n_x}^{(-)}\rangle$ with $x = n_x$, where $n_x \geq x_\rho$ [33]. In addition, all such ensembles have the same average size

$$\langle \text{MIC}_b \rangle_{n_x} = (q_{n_x}^{(+)} + q_{n_x}^{(-)}) \text{MIC}_b(n_x) = \frac{x_\rho - r}{n_x - r} \text{MIC}_b(n_x) \quad (\text{C6})$$

since the total weight

$$q_{n_x} \equiv q_{n_x}^{(+)} + q_{n_x}^{(-)} = \frac{x_\rho - r}{n_x - r} \quad (\text{C7})$$

only depends on n_x , see Fig. 5(c). Furthermore the case (i), discussed above, corresponds to the extremal case of $\langle \text{MIC}_b \rangle_{n_x=x_\rho}$ for which $q_{n_x} = 1$. Finally, note that from above n_x is bounded by $n_x \leq n_x^{\max}$ with

$$n_x^{\max}(x_\rho, z_\rho, r) = \begin{cases} 1 & \text{for } z_\rho \leq (1 - x_\rho) \sqrt{\frac{1+r}{1-r}} \\ \text{otherwise} & \frac{2r^2 x_\rho (x_\rho^2 + z_\rho^2 + 1) + r(2x_\rho^2(\sqrt{1-r^2}z_\rho - 3) + (z_\rho^2 - 1)(2\sqrt{1-r^2}z_\rho + z_\rho^2 + 1) - x_\rho^4) - 2x_\rho(x_\rho^2(\sqrt{1-r^2}z_\rho - 1) + (z_\rho^2 - 1)(\sqrt{1-r^2}z_\rho + 1))}{2z_\rho^2(2r^2 - 2rx_\rho + x_\rho^2 - 1) + (-2rx_\rho + x_\rho^2 + 1)^2 + z_\rho^4} \end{cases}, \quad (\text{C8})$$

as follows from a simple geometrical argument. This allows to write the size of ρ from Eq. (23) in the form

$$\widehat{\text{MIC}}_b(\rho) = \min_{n_x \in [x_\rho, n_x^{\max}]} \frac{x_\rho - r}{n_x - r} \text{MIC}_b(n_x) \quad (\text{C9})$$

with $r = I_0^{-1}(b)$, which is a well-behaved function that can be easily computed.

Appendix D: Weak Gaussian measurement for $I_\Delta(A : \ell)$

For the inequality (21), we calculate the relative entropy between an arbitrary distribution $p(x)$ and the Gaussian $g_\Delta(x - \bar{x})$, where $\bar{x} := \int dx xp(x)$. The relative entropy between two distributions $p(x), q(x)$ is defined as

$$S(p(x)||q(x)) := \int dx p(x) \log p(x) - p(x) \log q(x). \quad (\text{D1})$$

Thus,

$$\begin{aligned} S(p(x)||g_\Delta(x - \bar{x})) &= -H(p(x)) - \int dx p(x) \log g_\Delta(x - \bar{x}) \\ &= -H(p(x)) + \frac{1}{2} \log(2\pi\Delta^2) + \int dx p(x) \frac{(x - \bar{x})^2}{(2\ln 2)\Delta^2} \\ &= -H(p(x)) + \frac{1}{2} \log(2\pi\Delta^2) + \frac{\langle (x - \bar{x})^2 \rangle_{p(x)}}{(2\ln 2)\Delta^2}. \end{aligned} \quad (\text{D2})$$

Choosing $p(x) = \sum_\ell p_\ell g_\Delta(x - a_\ell)$ gives

$$\begin{aligned} S(p(x)||g_\Delta(x - \bar{x})) &= -H(p(x)) + \frac{1}{2} \log(2\pi\Delta^2) \\ &\quad + \frac{V(|\Psi\rangle, A) + \Delta^2}{(2\ln 2)\Delta^2} \\ &= -I_\Delta(A : \ell) + \frac{V(|\Psi\rangle, A)}{(2\ln 2)\Delta^2}, \end{aligned} \quad (\text{D3})$$

where we have used the fact that $p(x)$ is a convolution of p_ℓ and $g_\Delta(x)$, under which the variance is additive. The inequality then follows from the non-negativity of the relative entropy.

To show (22), we first prove the following useful result relating to classical statistics: Let A, X be random variables and $B := X + tA$. For sufficiently small t ,

$$H(B) = H(X) + \frac{t^2}{2\ln 2} \mathcal{F}_c(X) V(A) + O(t^3), \quad (\text{D4})$$

where $\mathcal{F}_c(X)$ is the classical Fisher information of X , and $V(A)$ is the variance of A .

Denote the density functions of X, B by $g(x), p(x)$ respectively, and let A have values a_ℓ with probabilities p_ℓ . The classical Fisher information of X is defined by

$$\mathcal{F}_c(X) := \int dx \frac{g'(x)^2}{g(x)}. \quad (\text{D5})$$

From the definition of B , we have

$$\begin{aligned} p(x) &= \sum_\ell p_\ell g(x - ta_\ell) \\ &\approx \sum_\ell p_\ell \left[g(x) - ta_\ell g'(x) + \frac{t^2 a_\ell^2}{2} g''(x) \right] \\ &= g(x) - t \langle A \rangle g'(x) + \frac{t^2 \langle A^2 \rangle}{2} g''(x), \end{aligned} \quad (\text{D6})$$

where we have done an expansion to $O(t^2)$. Similarly, it is easily shown that

$$\begin{aligned} p(x) \ln p(x) &\approx g(x) \ln g(x) + t [-\langle A \rangle g'(x) - \langle A \rangle g'(x) \ln g(x)] \\ &\quad + t^2 \left[\frac{\langle A \rangle^2 g'(x)^2}{2g(x)} + \frac{\langle A^2 \rangle}{2} g''(x) \right. \\ &\quad \left. + \frac{\langle A^2 \rangle}{2} g''(x) \ln g(x) \right]. \end{aligned} \quad (\text{D7})$$

In order to find $H(B)$, we integrate by parts

$$\begin{aligned} \int_{-\infty}^{\infty} dx g'(x) (1 + \ln g(x)) &= \left[g(x) (1 + \ln g(x)) \right]_{-\infty}^{\infty} \\ &\quad - \int_{-\infty}^{\infty} dx g'(x) = 0, \end{aligned} \quad (\text{D8})$$

$$\begin{aligned} \int_{-\infty}^{\infty} dx g''(x) (1 + \ln g(x)) &= \left[g'(x) (1 + \ln g(x)) \right]_{-\infty}^{\infty} \\ &\quad - \int_{-\infty}^{\infty} dx \frac{g'(x)^2}{g(x)} = -\mathcal{F}_c(X), \end{aligned} \quad (\text{D9})$$

assuming g is sufficiently regular that $\lim_{x \rightarrow \pm\infty} g'(x) \ln g(x) = 0$. Hence we have

$$-\int_{-\infty}^{\infty} dx p(x) \ln p(x) \approx -\int_{-\infty}^{\infty} dx g(x) \ln g(x) + \frac{t^2}{2} \mathcal{F}_c(X) (\langle A^2 \rangle - \langle A \rangle^2), \quad (\text{D10})$$

from which the result follows.

Now it can be verified that $I_\Delta(A : \ell)$ is unchanged under a simultaneous rescaling $\Delta \rightarrow \alpha\Delta$, $A \rightarrow \alpha A$. So $I_\Delta(A : \ell)$ in the limit of large Δ is the same as taking small t in $H(tA + X) - \frac{1}{2} \log(2\pi e/t^2)$, where $t = 1/\Delta$ and X is a standard Gaussian of unit variance and zero mean. Applying the above result, we get

$$I_\Delta(A : \ell) \approx \frac{t^2}{2 \ln 2} \mathcal{F}_c(X) V(|\Psi\rangle, A), \quad (\text{D11})$$

and $\mathcal{F}_c(X) = 1$.

Appendix E: Coherence criteria for $I_\Delta(A : \ell)$.

We need only check (C1) for pure states, as a mixed state has $I_\Delta(A : \ell)_\rho = 0$ if and only if there exists an ensemble decomposition with $I_\Delta(A : \ell)_{|\Psi_k\rangle} = 0 \forall k$. Now the concavity of the entropy tells us that $H(\sum_\ell p_\ell g_\Delta(x - a_\ell)) \geq \frac{1}{2} \log(2\pi e \Delta^2)$, with equality if and only if all the functions $g_\Delta(x - a_\ell)$ are the same, i.e. when $|\Psi\rangle = |A_\ell\rangle$ for some ℓ .

(C3) follows immediately from the convex roof definition.

(C2b) can first be shown for pure states. Suppose that a stochastic free operation takes $|\Psi\rangle \rightarrow |\Phi_\mu\rangle$ with probability w_μ , where $\sqrt{w_\mu} |\Phi_\mu\rangle = K_\mu |\Psi\rangle$, K_μ being a set of covariant Kraus operators, which take the form [18]

$$K_\mu = \sum_{i,j: a_i - a_j = \delta_\mu} c_{i,j}^\mu |A_i\rangle\langle A_j|. \quad (\text{E1})$$

Define p_ℓ as the probability of measuring $A = a_\ell$ for $|\Psi\rangle$, and p_ℓ^μ as the probability of $a_\ell + \delta_\mu$ for $|\Phi_\mu\rangle$. Then it can be shown that $p_\ell = \sum_\mu w_\mu p_\ell^\mu$ – this is because each p_ℓ^μ distribution is obtained by measuring and shifting p_ℓ by δ_μ . Let X be a standard Gaussian-distributed random variable. The distribution of $A + \Delta X$ for $|\Psi\rangle$ is $f(x) = \sum_\ell p_\ell g_\Delta(x - a_\ell)$, and similarly for $|\Phi_\mu\rangle$ we have $f_\mu(x) = \sum_\ell p_\ell^\mu g_\Delta(x - a_\ell)$. Now $f(x) = \sum_\mu w_\mu f_\mu(x)$ and so the concavity of the entropy gives $H(f) \geq \sum_\mu w_\mu H(f_\mu)$. It follows that $I_\Delta(A : \ell)_{|\Psi\rangle} \geq \sum_\mu w_\mu I_\Delta(A : \ell)_{|\Phi_\mu\rangle}$, as required.

(C2a,b) hold as a consequence of convexity and (C2b) for pure states – see [ref] for the argument.

Finally, for (C4), we let $|\Psi\rangle = (|A_i\rangle + |A_j\rangle)/\sqrt{2}$. It is clear that $I_\Delta(A : \ell)_{|\Psi\rangle}$ is a function of only $|a_i - a_j|$ and Δ . So the requirement that $I_\Delta(A : \ell)_{|\Psi\rangle}$ be increasing with $|a_i - a_j|$ is equivalent to it being increasing under a replacement $A \rightarrow \alpha A$ with $\alpha > 1$. By the scale-invariance property mentioned in Appendix D, this change of scale can be transferred to $\Delta \rightarrow \Delta/\alpha$. In fact, it should be clear that $I_\Delta(A : \ell)_{|\Psi\rangle}$ is decreasing with Δ (as more noise in the measurement cannot increase the mutual information) – so this property holds.

Appendix F: Coherence criteria for $\widehat{\text{MIC}}'_b$

For the vanishing criterion, note that either $I_0(A : \ell) \leq b$ or else there exists $\Delta > 0$ such that $I_\Delta(A : \ell) = b$ – this follows from it being decreasing with Δ and from $I_\Delta(A : \ell) \rightarrow 0$ as $\Delta \rightarrow \infty$.

(C2a) for $I_\Delta(A : \ell)$ immediately implies the same for $\widehat{\text{MIC}}'_b$. This is because $I_\Delta(A : \ell)_{\mathcal{E}(\rho)} \leq I_\Delta(A : \ell)_\rho$ (for covariant \mathcal{E}) shows that the set of values of Δ for which $I_\Delta(A : \ell)_{\mathcal{E}(\rho)} \geq b$ is at most as big as the corresponding set for ρ .

Similarly, since $I_\Delta(A : \ell)$ is an increasing function of $|a_i - a_j|$ for $|\Psi\rangle = (|A_i\rangle + |A_j\rangle)/\sqrt{2}$, the same is seen to be true for $\widehat{\text{MIC}}'_b$.

Appendix G: Coherence criteria for C_Δ

The channel taking induced by interacting M with P , and then tracing out P , is denoted by Φ^Δ – we refer to this as partial dephasing. Note that the full dephasing operation is $\mathcal{G} = \Phi^0$. We first list some useful properties of the partial dephasing channel:

- (i) $\Phi^\Delta = \int dk f_\Delta(k) \mathcal{U}_k$, where $f_\Delta(k) = \sqrt{\frac{2}{\pi}} \Delta e^{-\Delta^2 k^2}$ is the momentum-space probability distribution for $|\xi_\Delta\rangle$, and $\mathcal{U}_k = e^{-ikA}$;
- (ii) $[\mathcal{E}, \Phi^\Delta] = 0$ for any covariant channel \mathcal{E} ;
- (iii) $\Phi^\Delta(|A_i\rangle\langle A_j|) = e^{-\frac{(a_i - a_j)^2}{8\Delta^2}} |A_i\rangle\langle A_j|$;
- (iv) $\Phi^\alpha \circ \Phi^\beta = \Phi^\gamma$, where $\gamma^{-2} = \alpha^{-2} + \beta^{-2}$.

Proof. For (i), we perform the partial trace over P using momentum eigenstates $|k\rangle$:

$$\begin{aligned} \Phi^\Delta(\rho) &= \int dk \langle k|_P e^{-iA \otimes \hat{p}} (\rho \otimes |\xi_\Delta\rangle\langle\xi_\Delta|) e^{iA \otimes \hat{p}} |k\rangle_P \\ &= \int dk |\langle p|\xi_\Delta\rangle|^2 e^{-ikA} \rho e^{ikA} \\ &= \int dk f_\Delta(k) \mathcal{U}_k(\rho). \end{aligned} \quad (\text{G1})$$

(ii) then immediately follows, since $[\mathcal{E}, \mathcal{U}_k] = 0$ by definition. Instead tracing out P with position eigenstates, we have

$$\begin{aligned} \Phi^\Delta(|A_i\rangle\langle A_j|) &= \int dx \langle x|_P e^{-iA \otimes \hat{p}} (|A_i\rangle\langle A_j| \otimes |\xi_\Delta\rangle\langle\xi_\Delta|) e^{iA \otimes \hat{p}} |x\rangle_P \\ &= |A_i\rangle\langle A_j| \int dx \langle x| e^{-ia_i p} |\xi_\Delta\rangle\langle\xi_\Delta| e^{ia_j p} |x\rangle \\ &= |A_i\rangle\langle A_j| \int dx \langle x - a_i|\xi_\Delta\rangle\langle\xi_\Delta|x - a_j\rangle \\ &= |A_i\rangle\langle A_j| \int dx \frac{1}{\sqrt{2\pi}\Delta} e^{-\frac{1}{4\Delta^2}[(x-a_i)^2 + (x-a_j)^2]} \\ &= e^{-\frac{(a_i - a_j)^2}{8\Delta^2}} |A_i\rangle\langle A_j|, \end{aligned} \quad (\text{G2})$$

showing (iii); part (iv) follows from this expression. \square

In addition, C_Δ has the following properties:

- (a) $C_\Delta(\rho, A) \geq S(\rho || \Phi^{\Delta/\sqrt{2}}(\rho))$;
- (b) $C_\Delta(\rho, A) = \int dk f_\Delta(k) S(\mathcal{U}_k(\rho) || \Phi^\Delta(\rho))$;
- (c) decreasing with respect to Δ ;
- (d) invariant under a change of scale $A \rightarrow \alpha A, \Delta \rightarrow \alpha \Delta$.

Proof. (a) We need to use the fact that, for any quantum channel \mathcal{N} , $\text{tr}[\mathcal{N}(\rho) \log \sigma] \leq \text{tr}[\rho \log \mathcal{N}^\dagger(\sigma)]$, which is a consequence of the concavity of the logarithm [34] (Lemma 3.6). From this, we have

$$\begin{aligned} C_\Delta(\rho, A) &= -\text{tr}[\Phi^\Delta(\rho) \log \Phi^\Delta(\rho)] - S(\rho) \\ &\geq -\text{tr}[\rho \log \Phi^\Delta \circ \Phi^\Delta(\rho)] - S(\rho) \\ &= -\text{tr}[\rho \log \Phi^{\Delta/\sqrt{2}}(\rho)] - S(\rho) \\ &= S(\rho || \Phi^{\Delta/\sqrt{2}}(\rho)), \end{aligned} \quad (\text{G3})$$

having used property (iv) for the third line.

(b) From property (i) above,

$$\begin{aligned} C_\Delta(\rho, A) &= -\text{tr} \left[\int dk f_\Delta(k) \mathcal{U}_k(\rho) \log \Phi^\Delta(\rho) \right] - S(\rho) \\ &= \int dk f_\Delta(k) (-\text{tr}[\mathcal{U}_k(\rho) \log \Phi^\Delta(\rho)] - S(\rho)) \\ &= \int dk f_\Delta(k) (-\text{tr}[\mathcal{U}_k(\rho) \log \Phi^\Delta(\rho)] - S(\mathcal{U}_k(\rho))) \\ &= \int dk f_\Delta(k) S(\mathcal{U}_k(\rho) || \Phi^\Delta(\rho)), \end{aligned} \quad (\text{G4})$$

where for the third line, we have used the fact that \mathcal{U}_k leaves the entropy unchanged.

(c) Given some $\Delta_1 < \Delta_2$, there exists $\alpha \in (0, \infty)$ such that $\Delta_1^{-2} = \Delta_2^{-2} + \alpha^{-2}$. Then, by (iv) above, $\Phi^{\Delta_1} = \Phi^\alpha \circ \Phi^{\Delta_2}$. Therefore, taking d as the Hilbert space dimension,

$$\begin{aligned} S(\Phi^{\Delta_1}(\rho) || I/d) &= S(\Phi^\alpha \circ \Phi^{\Delta_2}(\rho) || \Phi^\alpha(I/d)) \\ &\leq S(\Phi^{\Delta_2}(\rho) || I/d), \end{aligned} \quad (\text{G5})$$

using the monotonicity of the relative entropy. Since $S(\rho || I/d) = \log d - S(\rho)$, this implies that $S(\Phi^{\Delta_1}(\rho)) \geq S(\Phi^{\Delta_2}(\rho))$. So C_Δ is decreasing with Δ .

(d) The scale-invariance is immediate from the fact that Φ^Δ multiplies the matrix element $|A_i\rangle\langle A_j|$ by a function of $(a_i - a_j)/\Delta$. \square

The properties of $\widetilde{\text{MIC}}_b$ follow exactly the same logic as for $\widetilde{\text{MIC}}'_b$ above.

Appendix H: Weak Gaussian measurement for C_Δ

We do the calculation for a general state with rank r strictly less than the dimension of the Hilbert space. We write the spectral decomposition $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, such that $\lambda_i > 0$ when $i < r$. Define the parameter $t = \Delta^{-2}$, which is assumed to be small. Let $\sigma_t = \Phi^{1/\sqrt{t}}(\rho) = \sum_i \mu_i |\phi_i\rangle\langle\phi_i|$, with its eigenvalues and eigenstates implicitly functions of t and coinciding with those for ρ at $t = 0$. We also write their t -derivatives at $t = 0$ as $\dot{\mu}_i, |\dot{\phi}_i\rangle$. To lowest order, $\mu_i \approx \lambda_i + t\dot{\mu}_i$, so

$$\begin{aligned} S(\sigma_t) &= -\sum_i \mu_i \log \mu_i \\ &\approx -\sum_i (\lambda_i + t\dot{\mu}_i) \log(\lambda_i + t\dot{\mu}_i) \\ &= -\sum_{i < r} (\lambda_i + t\dot{\mu}_i) [\log \lambda_i + \log(1 + t\dot{\mu}_i/\lambda_i)] \\ &\quad - \sum_{i \geq r} t\dot{\mu}_i [\log t + \log \dot{\mu}_i]. \end{aligned} \quad (\text{H1})$$

After the constant term $-\sum_i \lambda_i \log \lambda_i$, the leading order is $O(t \log t)$, so $S(\sigma_t) \approx S(\rho) - t \log t \sum_{i \geq r} \dot{\mu}_i$. Now,

$$\begin{aligned} \langle \psi_i | \partial_t \sigma_t | \psi_i \rangle &= \dot{\mu}_i + \lambda_i (\langle \psi_i | \dot{\phi}_i \rangle + \langle \dot{\phi}_i | \psi_i \rangle) \\ &= \dot{\mu}_i + \lambda_i \partial_t \langle \phi_i | \phi_i \rangle = \dot{\mu}_i. \end{aligned} \quad (\text{H2})$$

Since $\sum_i \mu_i = 1$ is constant, we have $\sum_{i \geq r} \dot{\mu}_i = -\sum_{i < r} \dot{\mu}_i = -\sum_{i < r} \langle \psi_i | \partial_t \sigma_t | \psi_i \rangle = -\text{tr}(P_\rho \partial_t \sigma_t)$.

It is easily verified that σ_t evolves according to the master equation

$$\partial_t \sigma_t = -\frac{1}{8} [A, [A, \sigma_t]]. \quad (\text{H3})$$

This can be seen by differentiating $\sigma_t = \Phi^{1/\sqrt{t}}(\rho)$ with respect to t , using property (iii) above of the dephasing channel. It follows that

$$\begin{aligned} \sum_{i \geq r} \dot{\mu}_i &= \frac{1}{8} \text{tr} (P_\rho [A^2 \rho + \rho A^2 - 2A\rho A]) \\ &= \frac{1}{4} \text{tr} (\rho A^2 - P_\rho A \rho A), \end{aligned} \quad (\text{H4})$$

having used the cyclic property of the trace and the fact that $P_\rho \rho = \rho P_\rho = \rho$. Putting these facts together gives the claimed result. For a pure state, $\rho = P_\rho = |\Psi\rangle\langle\Psi|$, so then $\text{tr}(P_\rho A \rho A) = \langle \Psi | A | \Psi \rangle^2$.

[1] E. Schrödinger, Die Naturwissenschaften **23**, 1 (1935).
[2] A. J. Leggett, Prog. Theor. Phys. Suppl. **69**, 80 (1980).
[3] A. J. Leggett, Journal of Physics: Condensed Matter **14**, R415 (2002).
[4] W. Dür, C. Simon, and J. I. Cirac, Phys. Rev. Lett. **89**, 210402 (2002).
[5] J. I. Korsbakken, K. B. Whaley, J. Dubois, and J. I. Cirac, Phys. Rev. A **75**, 042106 (2007).

[6] F. Marquardt, B. Abel, and J. von Delft, Phys. Rev. A **78**, 012109 (2008).
[7] P. Sekatski, N. Sangouard, and N. Gisin, Phys. Rev. A **89**, 012116 (2014).
[8] A. Laghaout, J. S. Neergaard-Nielsen, and U. L. Andersen, Optics Communications Macroscopic quantumness: theory and applications in optical sciences, **337**, 96 (2015).

- [9] See, for instance, the examples discussed in [5, 6].
- [10] G. Björk and P. G. L. Mana, *Journal of Optics B: Quantum and Semiclassical Optics* **6**, 429 (2004).
- [11] A. Shimizu and T. Morimae, *Phys. Rev. Lett.* **95**, 090401 (2005).
- [12] C.-W. Lee and H. Jeong, *Phys. Rev. Lett.* **106**, 220401 (2011).
- [13] F. Fröwis and W. Dür, *New Journal of Physics* **14**, 093039 (2012).
- [14] B. Yadin and V. Vedral, *Phys. Rev. A* **92**, 022356 (2015).
- [15] This is in analogy to the radioactive atom in the thought experiment of Schrödinger. Since N is large in general, note that the microscopic part is not literally expected to be a single atom with N levels. Rather, we might think of it as a small system composed of $O(\log(N))$ particles.
- [16] Our choice of the Shannon entropy here is not unique. One can easily think of contexts where other entropies would be more appropriate.
- [17] P. Sekatski, N. Gisin, and N. Sangouard, *Phys. Rev. Lett.* **113**, 090403 (2014).
- [18] B. Yadin and V. Vedral, *Phys. Rev. A* **93**, 022122 (2016).
- [19] S. L. Braunstein and C. M. Caves, *Physical Review Letters* **72**, 3439 (1994).
- [20] S. Yu, arXiv:1302.5311 (2013).
- [21] F. Fröwis, *J. Phys. A: Math. Theor.* (2017).
- [22] In contrast to $I(A : \ell)$, this quantum information is not necessarily completely extractable, due to discord. Nor is it necessarily information about ℓ , as the states $\{|A_\ell\rangle\}$ is only one particular basis choice for the system.
- [23] H. Ollivier and W. H. Zurek, *Physical review letters* **88**, 017901 (2001).
- [24] H. Kwon, C.-Y. Park, K. C. Tan, and H. Jeong, arXiv:1608.01122 [quant-ph] (2016).
- [25] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [26] The most direct way to argue is that $H(p(\ell|x))$ corresponds to the Von Neumann entropy of the partial state ρ_x^m after it undergoes a projection map in the basis $\{|\ell\rangle\}$, while the entropy is non-increasing under physical maps.
- [27] A. Winter and D. Yang, *Phys. Rev. Lett.* **116**, 120404 (2016), arXiv:arXiv:1506.07975v1.
- [28] E. Chitambar, A. Streltsov, S. Rana, M. N. Bera, G. Adesso, and M. Lewenstein, *Phys. Rev. Lett.* **116**, 1 (2016), arXiv:1507.08171.
- [29] A. Streltsov, E. Chitambar, S. Rana, M. N. Bera, A. Winter, and M. Lewenstein, *Phys. Rev. Lett.* **116**, 240405 (2016), arXiv:arXiv:1603.07508v1.
- [30] T. Baumgratz, M. Cramer, and M. Plenio, *Phys. Rev. Lett.* **113**, 140401 (2014).
- [31] L. P. Hughston, R. Jozsa, and W. K. Wootters, *Physics Letters A* **183**, 14 (1993).
- [32] G. M. D'Ariano, P. L. Presti, and P. Perinotti, *Journal of Physics A: Mathematical and General* **38**, 5979 (2005).
- [33] A careful reader might recall that four element ensembles on the plane are not extremal.
- [34] D. Sutter, M. Tomamichel, and A. W. Harrow, *IEEE Trans. Inf. Theory* **62**, 2907 (2016), arXiv:1507.00303.